# SRI International

Final Report • April 1997

# AUDITION AND VISION IN VIRTUAL ENVIRONMENTS

Prepared by:

Eugene Martinez-Uriegas, Principal Investigator
John D. Peters, Research Engineer
Nathaniel Bletter, Research Engineer
Visual Sciences Program
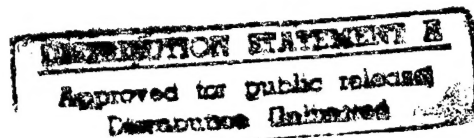Medical Technology Laboratory

SRI Project 5729

Prepared for:

Office of Naval Research
Ballston Tower One
800 North Quincy Street
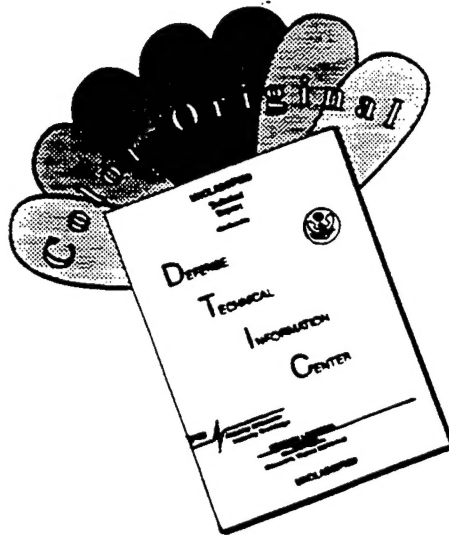Arlington, VA 22217-5660

Attention: Code ONR 342PS

Contract N00014-94-C-0097

19970508 041

# DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF COLOR PAGES WHICH DO NOT REPRODUCE LEGIBLY ON BLACK AND WHITE MICROFICHE.

# AUDITION AND VISION IN VIRTUAL ENVIRONMENTS

Prepared by:

Eugene Martinez-Uriegas, Principal Investigator
John D. Peters, Research Engineer
Nathaniel Bletter, Research Engineer
Visual Sciences Program
Medical Technology Laboratory

SRI Project 5729

Prepared for:

Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

Attention: Code ONR 342PS

Contract N00014-94-C-0097

Approved by:

Ajit Shah, Executive Director
Advanced Technologies Division

Harold E. Kruth, Senior Vice President
Advanced Technologies Division

# ACKNOWLEDGMENT

# CONTENTS

# ILLUSTRATIONS

# TABLES

# 1. SUMMARY

The scope of this study is the evaluation of the role played by simple orientation cues (visual and acoustic) within a virtual environment (VE). The main efforts during this project consisted of designing and conducting target-detection tests in a VE with and without visual and acoustic orientation cues. In addition, a system based on the SRI dual-Purkinje-image Eyetracker was set up to study eye movements induced by displaying moving target images at various image-update rates, as occurs in VEs. This report describes and analyzes experimental data on both visual and acoustic cues of orientation, as well as data on image-update rates and related eye movements.

The summary results of this project are

- Under the same conditions, target detection performance is better with a larger field of view in the VE. This result is consistent with many measurements of task performance within real environments.

- Unlike results obtained within real environments, three-dimensional visual orientation grids in a VE can be detrimental for target detection tasks under certain conditions.

- Adding cues to a three-dimensional orientation grid in a VE, thereby providing absolute direction of orientation (north, south, east, west), does not improve target detection performance.

- When an acoustic compass is added to relative or absolute visual orientation grids in a VE, target detection performance is reduced. However, adding an acoustic cue attached to the target object in a VE significantly improves target detection performance.

- Update rates of the position of a moving target in VEs are linked to saccadic and smooth-pursuit eye movements. Below 5 Hz, eye movements are predominantly saccadic; above 15 Hz, eye movements are predominantly smooth-pursuit. These two types of eye movement rapidly alternate for intermediate update rates as measured by dynamic fixation error and average eye speed.

Summary guidelines for VE designers derived from these results are

- Improving resolution in head-mounted displays by reducing field of view should be avoided because it severely reduces the user's performance of visual tasks.

- When a designer is planning to increase the field of view of a VE, not only should he take measures to maintain spatial resolution, but also consider a corresponding increase in image update rate to avoid reductions in visual performance. Image update rate is more important for large fields of view, 40 degrees and larger, where, according to the data of this study, an image update rate of the order of 16 Hz is linked to considerable reductions in visual performance.

- The use of acoustic cues similar to an acoustic compass is not recommended for VEs because it does not help spatial orientation and introduces unnecessary distraction. However, it is of great benefit to attach an acoustic cue to a virtual object that changes spatial location within the VE, because it increases the user's efficiency to search, find, and identify the object.

- It is recommended to avoid image update rates within the 5- to 15-Hz range because they are a factor of negative psychophysiological effects in the user, especially with fields of view 40 degrees and larger.

- For VEs that are delivering position update rates within 5 to 15 Hz, it is recommended to train the user to avoid the rapid alternation of saccadic and smooth-pursuit eye movements and voluntarily restrict his eye movements to just smooth pursuit when the optical flow of the VE is jerky and discontinuous.

## 2. VIRTUAL ENVIRONMENT SETUP

In this series of experiments, subjects are required to locate a specific target within a VE that may or may not contain visual and/or acoustic orientation cues. The virtual environment consists of a uniform, white virtual sphere with a diameter of 40 feet, upon which an optional spherical grid of meridian and parallel lines can be superimposed. Figure 1 is a line drawing of the subject and the virtual sphere. There are two types of optional grid, the relative orientation grid which comprises just meridian and parallel lines, and the absolute orientation grid which has the initials N, S, E, and W placed along the corresponding meridian lines. Each initial is displayed several times along its corresponding meridian line such that the subject will always see at least one of the letters independently of the elevation of his gaze. Acoustic cues are spatially localized sounds within the VE. For clarity, they are separately described in Section 3. The meridian lines are placed at 30-degree intervals, and the parallel lines are placed at the equator and at ±15-degree increments up to 45 degrees from the equator. The target and distractors are randomly presented anywhere within ±45 degrees from the equator, and at any heading.



(a)                    (b)                    (c)
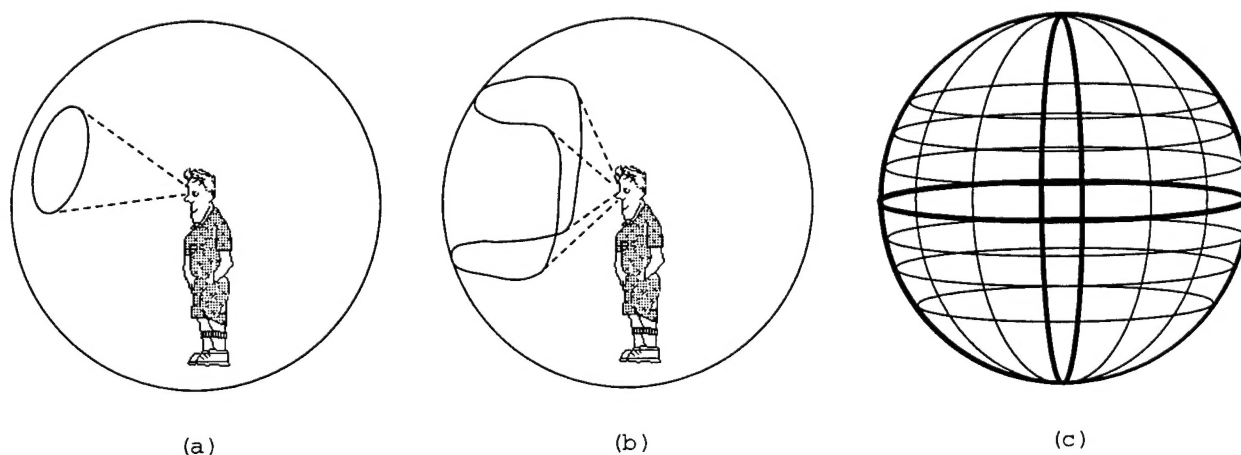
Figure 1.   Sketch of the virtual environment. The drawings show the spherical wall of the virtual environment. (a) The subject's field of view is limited within circular apertures of 14, 28, 41, and 53 degrees of visual angle. (b) Illustration of the widest field of view, 60 by 100 degrees, equivalent in cross-sectional area to a circular aperture of 87 degrees. (c) Illustration of the orientation grid used inside the virtual sphere.

The target is always a red square subtending about 3.3 degrees (200 minutes) of visual angle per side. Distractors consist of four red discs and four green squares, all subtending the same visual angle as the target. In addition to the wide field of view condition, where the subject views the entire 60-degree-vertical by 100-degree-horizontal image of the virtual sphere, testing

was carried out with circular visual apertures of 14, 28, 41, and 53 degrees of visual angle. These apertures have virtual edges and background that perfectly match the virtual sphere. Thus, the subject cannot tell that he is viewing the virtual sphere through an aperture unless virtual features, such as meridian and parallel lines of the grid, or the target or distractors, happen to be partially occluded by the aperture's edge. For comparison purposes, the full (rectangular) visual field of 60 by 100 degrees will be considered here to be an 87-degree circular aperture, just on the basis of having the same cross-section area. With this clarification, the fields of view tested in this project have apertures of 14, 28, 41, 53, and 87 degrees of visual angle.

The virtual sphere and all virtual features are produced on the binocular liquid-crystal displays (LCDs) of a set of VPL EyePhones.[1] Pixel pitch of the LEEP-Optics-magnified images of the LCDs is approximately 13 minutes of visual angle. Thus, the upper limit of spatial frequencies that can be presented in the EyePhones is approximately two cycles per degree of visual angle.

Head tracking is accomplished with a Polhemus magnetic tracker that samples head position and orientation at 60 Hz. Tracker range is accurate only over a 3-foot radius, so the subject is constrained to this volume. To aid the subject, a 5-inch plastic disc is affixed to the floor directly under the Polhemus emitter. The subject is instructed to maintain his position within the 3-foot radius from the emitter by feeling the disc with his foot. Pitch and yaw are measured and recorded at 60 Hz.

The subject wears a Cyber Glove to communicate his responses to the computer controlling the experiment. By clenching his fist, the subject signals that he has found the target. The computer verifies that the target is currently visible within the preset aperture and records the response as a hit; otherwise, the response is rejected.

The VE system runs on two Silicon Graphics R4000-50 VGX (one for each display of the EyePhones, plus the Cyber Glove and the Polhemus head tracker), an IBM PC runs the Convolvotron, and all are linked by a fourth machine (a Macintosh IIfx).

Two performance units were initially adopted for these tests, the inverses of search time and visual pathlength. Search time was recorded in head-tracker sampling intervals, that is, every 1/60 of a second. Pathlength was calculated as the sum of the number of degrees of pitch and the number of degrees of yaw between successive hits. Thus, long detection times or large pathlength values correspond to low performance, and vice versa. Pathlength was selected as an alternative metric in these experiments because pilot studies indicated that the presence of visual orientation cues (grids) in the virtual sphere caused some subjects to slow their head movements, ostensibly to reduce vertigo (an important observation that influenced the decision of conducting eye movement measurements in the last part of this project). During an experimental session, total search time between successive hits and total pathlength between successive hits were recorded for 32 consecutive presentations. Only the last 30 responses were actually registered and analyzed, in order to avoid initialization response biases.

---

[1] All product names mentioned in this report are the trademarks of their respective holders.

## 2.1    RELATIONSHIP BETWEEN FIELD OF VIEW AND PERFORMANCE

In similarity with object identification in real environments [1, 2], target-detection performance in a VE, with and without visual orientation cues, shows a clear first-order relationship with field-of-view size.
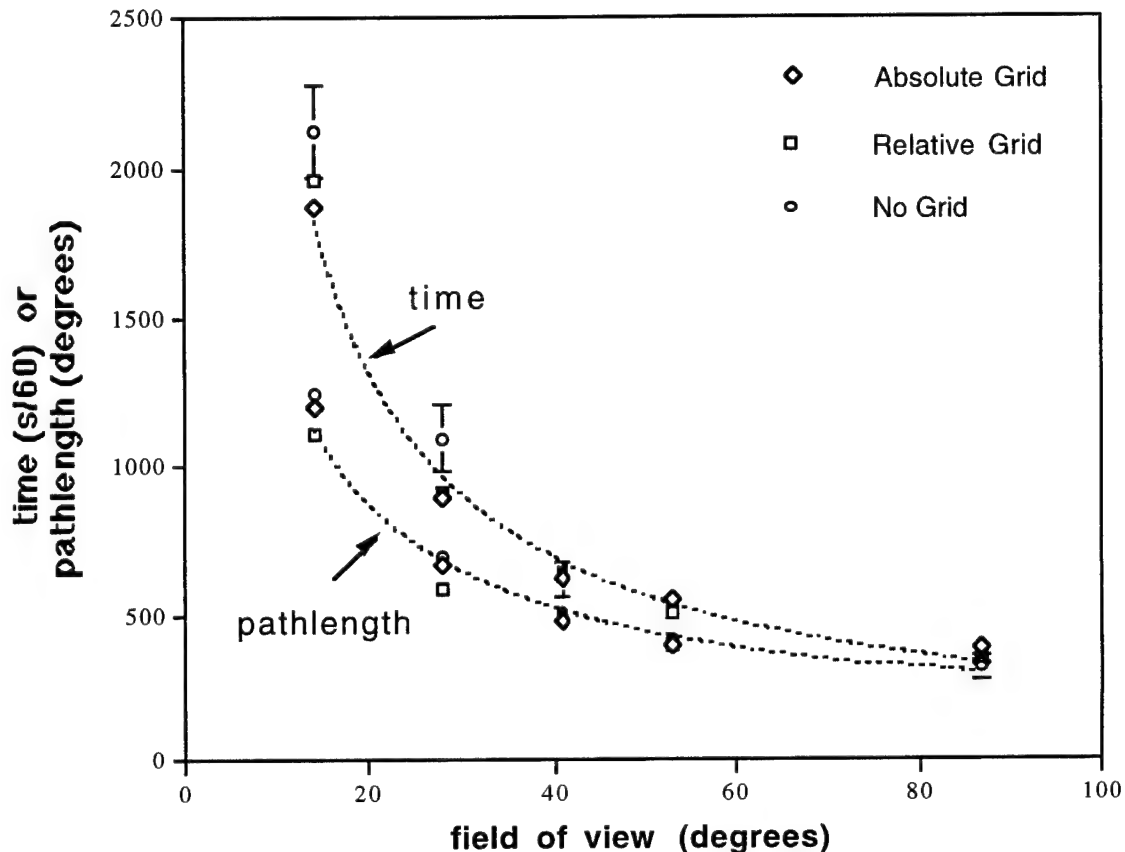


Figure 2.    Search time and pathlength versus field of view. Average results from fourteen subjects show the search time and the pathlength required for target detection (ordinate) as functions of the field of view of the subject (abscissa). The dashed lines are hand drawn to emphasize the separate clustering of the two types of measures.

Data plotted on Figure 2 are average results from fourteen subjects tested with five different apertures restricting the field of view, and each under three conditions: (1) the basic VE without any visual orientation cues, where the subject has the task of detecting a target in an empty, uniform white environment, and where the only objects are the target and distractors appearing at random places on each trial, (2) the same VE but with the addition of a relative orientation grid of parallel and meridian lines attached to the environment, and (3) with additional cues placed on the lines of the grid to identify absolute direction (north, south, east, west).

5

The graph shows how performance varies directly with field size. Measurements with both pathlength and search time show that performance is quite poor with small fields of view (i.e., pathlength and search time have large values), improves significantly with field sizes exceeding about 30 degrees, and begins to asymptote at the 87-degree field of view where the highest performance values (i.e., the lowest values of pathlength and search time) are obtained.

The two main clusters of data in Figure 2 correspond to the two different ways of measuring performance, search time and pathlength, and they have roughly the same shape which is an indication that the two methods are probably measuring the same property. Both sets of data are fitted reasonably well by power functions (all individual curves fit a log-log relationship with a goodness of fit above 0.9). However, search time is preferable because (1) time measurements span a broader range than pathlength (i.e., a scale of higher precision), and (2) the time measurement (in seconds/60) is directly measured by the computer clock between the start and the end of the trial, while the pathlength measurement (the addition of sequential yaw and pitch displacements during a trial) depends on the inherent lower precision of the head tracker, and individual displacement measurement errors accumulate in the addition of consecutive head positions every sixtieth of a second until the target is detected. Consequently, only the performance units based on search time will be used hereafter in this report. Table 1 contains summary data of search time under all testing conditions with visual and acoustic orientation cues. The entries are averages of all fourteen subjects participating in this study. The results with different acoustic cues are discussed in Section 3.

Figure 3 shows the log-log fit of search time data. In all three conditions, performance increases with increasing field of view. Nonetheless, it is interesting that for small fields of view, the condition with absolute orientation cues gives the best performance and the condition without any orientation cues has the lowest performance, while the converse is true for large fields of view. This is discussed in detail in the next section.

For this part, the result of increased performance for increased field of view in VEs suggests that one of the current methods of improving resolution in head-mounted displays—by reducing field of view—severely reduces the user's ability to find objects in a VE. The performance curves indicate that field of view in VEs, whether through a head-mounted display or otherwise, should probably be 30 degrees of visual angle or better, so that the user may perform object detection and identification tasks with reasonable efficiency.

## 2.2 EFFECTS OF VISUAL ORIENTATION CUES ON TARGET DETECTION

When the user of a VE is given the task of searching for a specific object that is present somewhere in a 360-degree section of a sphere around him, it seems natural to assume that any cue providing information on where he has already looked should help him to conduct the task more efficiently. That should be even more important when the field of view is restricted, requiring many more visual sweeps, head movements, and body postural changes to cover the surrounding space, thereby making it harder to keep track of the regions already explored.

**Table 1**

**MEASUREMENTS OF TIME REQUIRED FOR TARGET DETECTION**

Search time values (the units are in 1/60 second) for nine sets of visual and acoustic testing conditions. Each set was tested with five different fields of view given in the first column as aperture values (the units are degrees of visual angle). Data are averages from fourteen subjects and are plotted in Figures 1 and 2.

| aperture | no grid no sound | absolute grid no sound | relative grid no sound |
|---|---|---|---|
| 14.00 | 2127.71 | 1875.00 | 1967.43 |
| 28.00 | 1095.43 | 894.57 | 910.79 |
| 41.00 | 620.00 | 620.00 | 638.71 |
| 53.00 | 521.07 | 551.36 | 507.79 |
| 87.00 | 320.50 | 392.07 | 378.64 |
| aperture | no grid compass sound | absolute grid compass sound | relative grid compass sound |
| 14.00 | 2520.24 | 2088.00 | 2279.28 |
| 28.00 | 1289.53 | 994.57 | 1044.51 |
| 41.00 | 737.34 | 672.46 | 720.91 |
| 53.00 | 617.65 | 602.38 | 570.03 |
| 87.00 | 370.80 | 430.46 | 410.23 |
| aperture | no grid target sound | absolute grid target sound | relative grid target sound |
| 14.00 | 1480.32 | 1050.76 | 1220.08 |
| 28.00 | 831.30 | 591.11 | 560.93 |
| 41.00 | 473.23 | 440.24 | 370.29 |
| 53.00 | 408.16 | 396.52 | 320.28 |
| 87.00 | 253.92 | 280.27 | 250.64 |

Such common sense reasoning, generally true for real environments [1], was found to hold only for small visual fields of the VE but not for the testing conditions with larger visual fields. Figure 4 shows average results from fourteen subjects on the effects of visual orientation cues upon target detection in a VE with different fields of view. The data on the ordinate are the relative changes in performance with respect to the condition without any orientation cues. Change in performance due to the addition of orientation cues was computed from search time data using a normalized ratio where search time is inversely proportional to performance. Let $T_r$ and $T_n$ be the search times with relative orientation cues and no orientation cues, respectively; then, the relative change in performance ($\Delta_{pr}$) due to the addition of relative orientation grid is

$$\Delta_{pr} = (T_n - T_r) / T_n \tag{1}$$

and similarly for the relative change in performance ($\Delta_{pra}$) due to the addition of absolute orientation cues (upon the orientation grid):
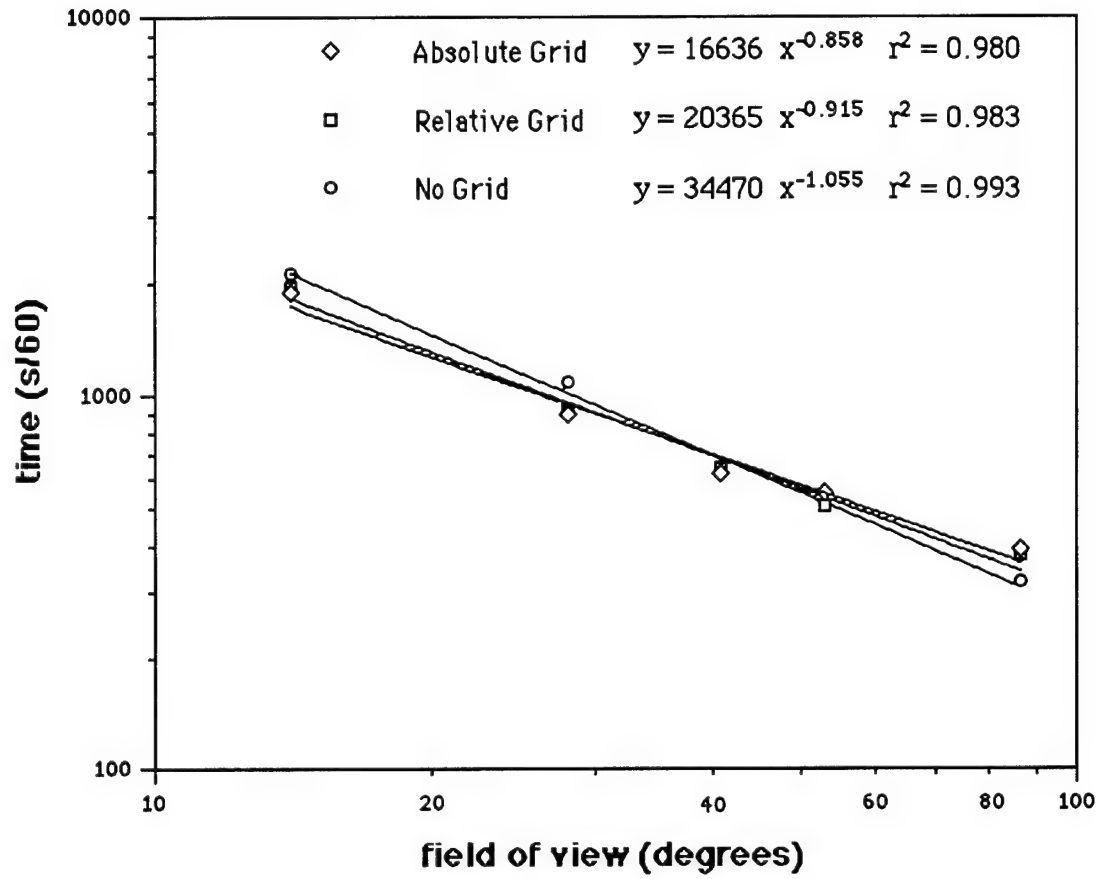
$$\Delta_{pra} = (T_n - T_a) / T_n \tag{2}$$

Figure 3. Search time versus field of view in log-log coordinates. Summary results from the three orientation conditions are fit by power functions, as indicated. Note the differences in slope.
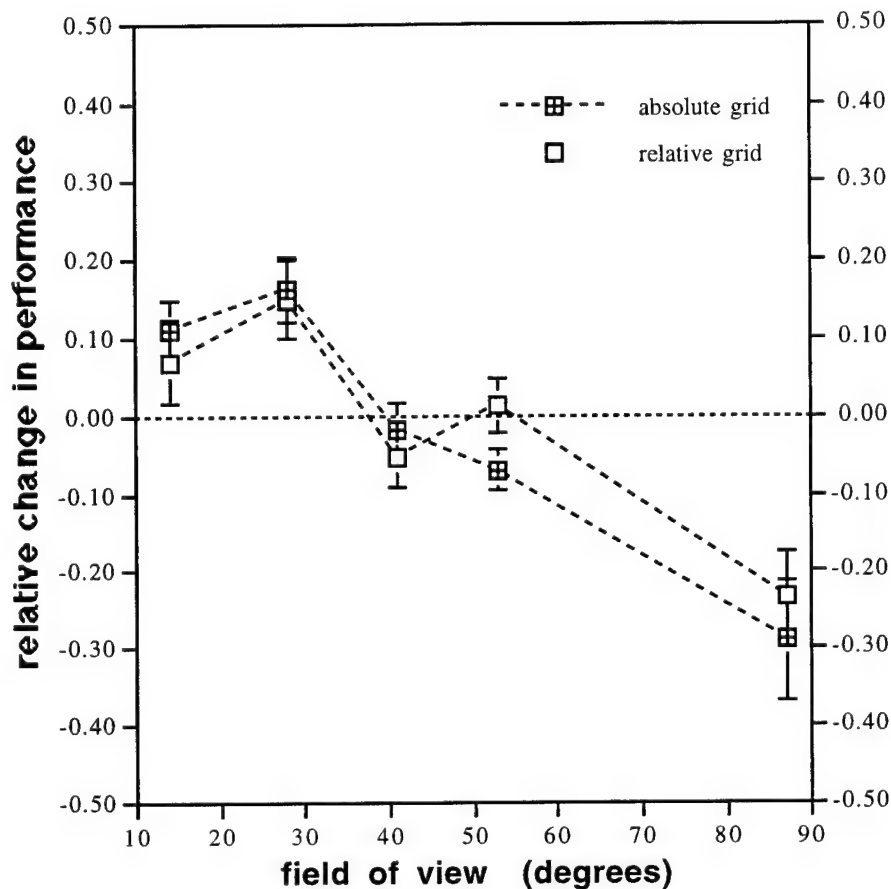
8

Figure 4.    Summary plot of relative and absolute orientation cues. Comparison plot of absolute and relative cues in visual orientation grids in terms of changes in performance (ordinate) for different fields of view (abscissa). The reference level (dashed line at zero) corresponds to the performance without any visual orientation cues in the VE.

where $T_a$ is the search time measured when absolute orientation cues are added to the orientation grid in the VE.

The data show that, for a small field of view of 28 degrees, a relative orientation grid increases target detection performance by 15%; adding absolute orientation cues does not seem to make a significant difference (total 16% increase in performance). However, for the largest field of view of 87 degrees, the orientation grid *reduces* target detection performance by 23%; introducing the absolute orientation cues on the grid makes it worse, for a total reduction of 29% in performance.

Figure 5 shows the differential effect of adding the absolute orientation cues; it is simply the difference of the two curves in Figure 4. Although there seems to be a trend where absolute cues help performance in smaller fields of view and hurt performance in the larger fields of view, the experimental error bars indicate that this effect is not significant, except at the widest
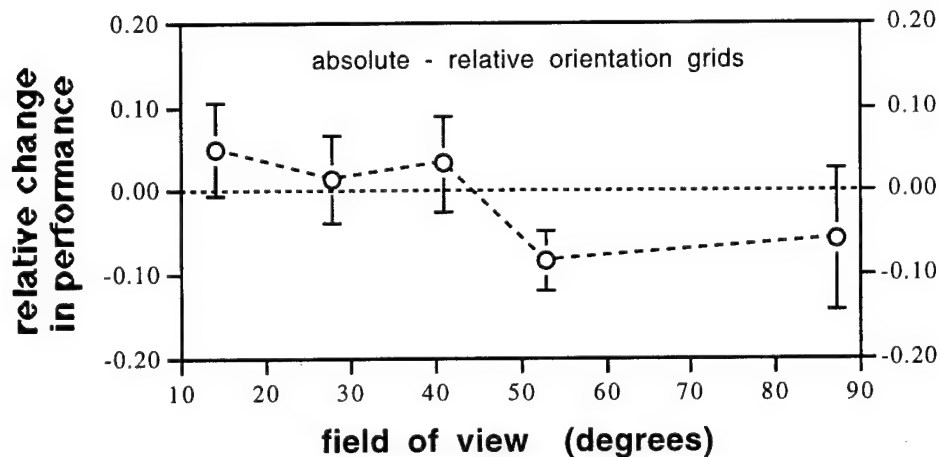
Figure 5.    Differential effects of absolute and relative orientation cues. Plot of the difference of data pairs from Figure 4 for each field of view.

field of view (87 degrees). In early experiments during this study the absolute orientation cues were not the direction letters N, S, E, and W. Instead, the corresponding lines on the grid were color coded red, green, yellow, and blue. In those conditions there was no observed effect on performance from adding the absolute orientation cues. The possible explanation for that was that the red color cue for absolute direction was interfering with the color cue for the target (red), introducing additional difficulty for the detection task, such that it counterbalanced the possible advantages introduced by the absolute orientation cues, and that was the reason to use letters instead. The results shown in Figure 5 indicate that the possibility of color cue interference should be rejected because the results are similar without the use of color coding, unless it could be argued that the letters produce an interference effect similar to that of color, but that does not seem plausible unless it could be demonstrated that both color and letter cues play a similar role in the demand of attention from the subject while he is performing the search task, and that is a matter for yet another study.

The results of reduced performance for the largest field of view shown in Figure 4 are somewhat surprising because in real environments visual orientation cues usually improve performance of target search and detection, even when the scene is viewed through optical stereoscopic devices [3]. As described earlier, the orientation grid is made of parallel and meridian lines. This means that for each size of field of view there is a different number and length of visible grid lines within the visual aperture. When the subject moves his head to search for the target, all the visible lines of the grid within the aperture of the field of view move simultaneously and in the opposite direction of the subject's head movement. In real environments, this optical flow occurs in real time and is perceived as a smooth motion synchronized with the signals from the semicircular canals about head position. In virtual environments, the simulated optical flow depends on the position update rate of the imaging system which, in most currently available apparatus including the one used for this study, is not

fast enough to be perceived as a smooth motion flow and is not in synchrony with the head position perception of the subject. With the system used for this study, and with the specific three-dimensional environment that included the orientation grid, the update rate of the stereo imaging system was about 16 Hz. The percept of the subject is that the grid lines shift through the field of view in a jerky fashion, reportedly producing in some cases a sensation of body unbalance and in some cases vertigo, inducing the subject to intuitively reduce the speed of head movements (thereby increasing search time) to reduce the apparent asynchrony. With a necessary remark for caution because of the subjective nature of verbal reports and in this case only by about half of the subjects, sensations of body unbalance are not perceived for the smaller fields of view of 14 and 28 degrees, appear for the 41-degree field of view, and are increasingly conspicuous and stronger for the larger fields of view of 53 and 87 degrees. This reasoning leads to the speculation, based also in previous reports [4], that the presence and strength of these sensations is related to the density and angular extent of grid lines visible within the field of view and shifting simultaneously in a jerky fashion. Additional studies with focus on those variables are needed to quantify these suggested correlations.

In any case, there are important practical implications from the finding that, under the stated conditions, visual orientation cues are detrimental for visual detection tasks in VEs. Currently, one of the main interests of VE designers is to increase the image update rate of virtual imaging systems to produce a smoother motion of simulated optical flow, thereby increasing the degree of perceived immersion in the virtual scene. The results of this study indicate the additional consideration that the image update rate is much more critical for large fields of view, 40 degrees and larger, where an image update rate of the order of 16 Hz (the average rate delivered by our laboratory setup) is linked here to considerable reductions in visual performance. In more practical terms, when a designer is planning to increase the field of view of a VE, not only should he take measures to maintain spatial resolution, but also consider a corresponding increase in image update rate to avoid reductions in visual search performance.

# 3. ACOUSTIC ORIENTATION SETUP

These studies test the effectiveness of acoustic sources as orientation cues in a VE [5,6]. The experimental tests were conducted using the same apparatus as in the case of the visual orientation cues, but adding virtual acoustic icons into the subject's visual world.

Two types of acoustic icons were selected. One type is an acoustic compass: a characteristic sound that appears to be emitted always from the same point (from the north direction at the equator of the virtual sphere), independently of any motion of objects in the scene and any motion of the subject. The sound is easily discriminable and intermittent (a continuous sound was tested and rejected because it annoyed the subjects and was harder to locate in space). The acoustic compass is the auditory equivalent of the absolute visual grid, although it was simplified from four different sounds (used in initial experiments) to just one, because the four sounds introduced unnecessary redundancy and distraction of the subject. The subjective experience of the sound compass reported by the subjects is that they can easily identify the sound and establish stationary directions of "north, south, east, and west". However, making a decision about his own orientation in the virtual scene at a given moment usually requires the subject to pay special attention to that sound while swaying his head to zero-in on the apparent direction of the acoustic source.

The second type of auditory cue is a target-specific acoustic cue, which is always attached to the visual target and is perceived as emanating from whatever location in the virtual scene is occupied by the target. The reported subjective experience is that this cue becomes an additional property of the target itself. This icon is the auditory equivalent of the target's visual characteristics, adding to the target-specific combination of shape and color (red square), a target-specific sound.

All acoustic cues are generated by a Crystal River Convolvotron, which creates sounds that appear to be localized in three-dimensional space. It essentially recreates the transit delays and head-related transfer functions that the auditory system uses to localize sound sources.

## 3.1    EFFECTS OF ACOUSTIC ORIENTATION CUES

The results in Figure 6 show three main findings: (1) there is a significant improvement in target detection performance, 25% to 40%, when the target-specific acoustic cue is added to the VE under any combination of other stimulus parameters; (2) the acoustic compass has the opposite effect; it produces reductions of about 15% in performance for all testing conditions; and (3) the field of view size does not seem to have a significant effect; that is, both the improvement in performance with the target sound and the performance decrease with the compass sound are practically independent of the size of field of view.
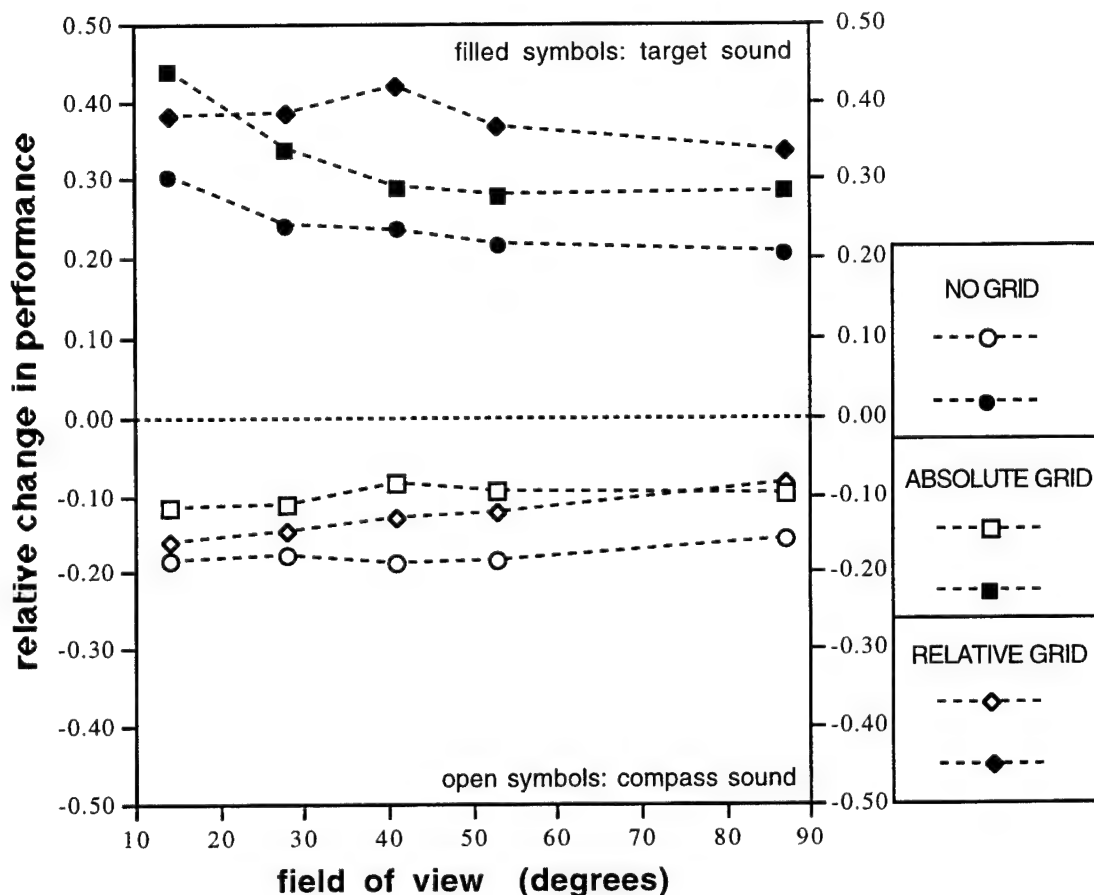
12

Figure 6.    Summary plot of acoustic compass and target sound cues. Average results of changes in performance from fourteen subjects under all the tested combinations of visual and acoustic cues for orientation. The reference level (dashed line at zero) corresponds to performance under the same conditions but without any acoustic cues.

Figure 6 is the summary plot of all acoustic conditions normalized and averaged across all subjects. The values of change in performance are with respect to the corresponding condition without any acoustic cues and are computed with a formula similar to the one used for visual cues:

$$\Delta_p = (T_o - T_s) / T_o \qquad (3)$$

where $\Delta_p$ is the change in performance, and $T_o$ and $T_s$ are the measured search times without the sound cue and with the sound cue, respectively. For example, a change in performance of +0.25 means that the subject required 25% less time to detect the target when the acoustic cue was present than he did without the acoustic cue.

A possible interpretation of these results in terms of cognitive space is as follows. Visual and acoustic orientation cues provide elements to build an internal representation of space that is used by the subject to perform target localization and identification tasks.

When only the target and distractors are present, the internal representation lacks reference landmarks because both target and distractors randomly change position every trial. Nonetheless, the internal representation acquires the characteristic of an approximately spherical surrounding from the fact that target and distractors always appear at the same stereoscopic distance away from the subject in VE.

When the orientation grid is introduced in the VE, the internal representation acquires a set of structural landmarks that more accurately define the spherical surrounding. The upper and lower boundaries of the grid define the upper and lower parallels at $\pm 45$ degrees beyond which target and distractors never appear. The grid is more useful with smaller fields of view because it provides references to scan the environment with more efficient coverage and fewer sweeps. Higher performance figures were obtained in these conditions. However, when the field of view increases over 40 degrees, larger sections of the grid are viewed simultaneously, and when the subject moves his head, they appear to move in a jerky fashion that sometimes produces body unbalance. It is probable that the striking increase in perceived "jerkiness" of the grid's motion for larger fields of view is due to the participation of peripheral visual mechanisms, which are more sensitive to temporal changes than the mechanisms corresponding to the central retina. Correspondingly, the performance measured for large fields of view is low, even below levels measured without the orientation grid.

The introduction of an acoustic compass provides information about absolute orientation, acting as an anchor for the internal representation of space, but our results show that it makes more difficult the task of target search and detection. One possible explanation for this unexpected result is the apparent attention of the subject required by the auditory system to localize and keep track of the position of a sound source and integrate that spatial information dynamically into the internal representation of space [6]. That is, the task of incorporating and updating that information into the internal representation of space seems to be a difficult task in itself, perhaps even more difficult than searching visually for the target, and it consumes valuable attention resources that can be more efficiently used to visually find the target.

On the other hand, the target-specific acoustic cue improves performance in all measured conditions. This acoustic cue provides additional information about the target position because the Convolvotron uses the target position to simulate the spatial location of the source of this sound, so the sound always appears to originate from the target. It does not contribute to the internal representation of space but makes the task easier because it signals the general direction of localization of the target when it is out of the visual field; the subject can then turn his head in that general direction and find the target visually, without having to inspect many other possible directions. This is possible because the acoustic cues used in this project are simulated on a full-surround uniform space; they do not have the equivalent of the virtual apertures used to restrict the field of view. Thus, even if the subject is working under, say, a 14-degree field of view, the acoustic cue attached to the target is propagated throughout the full surround, providing positional information that is not available through the small visual aperture.

That is also the reason for the relative independence of the acoustic measurements with respect to the field of view. All acoustic cues are projected by the Convolvotron from the source point in all directions, without any acoustic apertures.

In summary, the results of tests with acoustic cues for spatial orientation provide two guidelines: (1) the use of acoustic cues similar to an acoustic compass is not recommended for virtual environments where the user is expected to visually search, detect, and identify certain objects with the best possible efficacy; and (2) an acoustic cue attached to a virtual object that changes spatial location within the VE increases the user's efficacy to search, find, and identify the object.

# 4. UPDATE RATE AND EYE MOVEMENTS

The objective of these experiments is the measurement of eye movements of subjects who are asked to visually track a moving target. The position of the target on a screen is updated at selected rates by the experimenter. The motivation for these tests in the context of virtual environments comes from the finding discussed in Section 2 that, under certain conditions, target detection performance in VEs is reduced by the addition of orientation grids, while similar orientation cues improve performance in real environments.

It is well known that one of the pervasive technical problems in current VEs is the delay between the subject's head movement and the corresponding update of the position of virtual objects, mostly because of the complexity and time requirements of high-resolution three-dimensional graphics rendering. These delays not only reduce the subject's experience of immersion, they even produce motion sickness and vertigo in some users. It is possible that the reductions in performance reported in Section 2 are due to the delay in position update of the orientation grid. This delay could be producing certain patterns of eye movements which send neural signals that are in conflict with body-position signals from the semicircular canals [7]. The subject may be intuitively reducing the speed of his head and body movements to reduce the conflict of signals, and that could be the origin of the measurements of longer search time.

The results of accurately monitoring eye movements discussed below show that, at high position update rates of the target, such as 20 or 30 Hz, the subject makes smooth-pursuit eye movements to follow the moving target with high accuracy, while at very low update rates—1 to 2 Hz, for example—the subject makes saccadic eye movements almost exclusively and his fixation on the target is less accurate. Intermediate conditions show how accurately the subject is tracking the stimulus under different update rates and also evaluate the relative predominance of saccadic or smooth-pursuit eye movements involved in that task.

These tests provide an accurate comparison between the actual trajectory of the target and the trajectory of the eye fixation point. From these measurements is derived the absolute error between the actual position of the target and the fixation position of the subject's eye, as a function of target position update rate. Another variable of interest is the average velocity of the subject's eye under different position update rates.

## 4.1    APPARATUS

The computer-controlled apparatus consists of three subsystems:

- A monitor displaying a computer-generated target that moves at selected average speeds and selected update rates.

- An SRI dual-Purkinje-image Eyetracker that samples the subject's eye position and rotation at several kilohertz with an accuracy of approximately 1 minute of arc.

- A MacADIOS interface that samples (at 100 Hz) and digitizes the two-dimensional eyetracker signals; it also receives in synchrony the stimulus parameters and processes the data.

## 4.2    EXPERIMENTAL CONDITIONS AND PROCEDURE

In a first set of experiments the target was designed to follow a back-and-forth horizontal straight path. The analysis of results showed a bias in the measurements, because the eye's point of fixation at the end points of the trajectory was unusually far from the actual target position. It was either overshooting or anticipating the actual turning point, producing a larger absolute error in both cases, and that was not related to the position update rate but just to the sudden reversal of direction of motion. Therefore, a second set of experiments was necessary, and a circular path was used instead.

The monitor is placed at a distance of 122 centimeters from the subject; the screen subtends 18.4 degrees of visual angle at that distance. A square target of controllable size and luminous contrast moves on the screen of the monitor at a fixed speed preset by the experimenter. Several pilot tests led to the selection of a target consisting of a 0.14-degree square with 75% luminance contrast moving at an average speed of 85.7 degrees per second (about 4.2 seconds per revolution) along a circular path with diameter subtending 9.6 degrees of visual angle. With these target parameters fixed, measurements were recorded of the eye movements of three subjects (naive with respect to the details and purpose of the experiments) for twelve different values of position update rate of the target.

A position update rate of 1 Hz, for example, causes the position of the target to be changed only four times during a full cycle, while at 30 Hz the position changes about 126 times per cycle. At 1 Hz, the subject's eye movements are predominantly of the saccadic type, while at 30 Hz they are predominantly smooth pursuit movements. In both cases the average speed of the target (angular distance covered between position updates divided by the time between updates) is the same because it is maintained at 85.7 degrees per second, but the average speed of the eye is expected to be higher for saccadic movements than for smooth pursuits. With respect to the fixation error—that is, the difference between the actual position of the target and the subject's fixation point—higher error values are expected for update rates where saccadic eye movements predominate—that is, at low update rates, while low error values should result at high update rates where smooth-pursuit eye movements predominate.

The subject sits in front of the computer monitor with his head immobilized by using a bite bar with his dental impression. The bite bar is firmly attached to the same bench where the eyetracker is installed. This is necessary to avoid the combination of head and eye movements which would make practically impossible an adequate control of variables and would prevent accurate measurements. After calibration (horizontal and vertical), the eyetracker monitors the subject's eye movements as he visually tracks the moving target during 20 seconds for each of twelve different position update rates: 1, 2, 4, 6, 7.5, 8.57, 10, 12, 15, 20, 30, and 60 Hz. There is

17

an important point relating these tests with VEs. In these measurements, the target is being moved by changing its position at a given update rate by the computer program. In a VE, data about the motion of the subject's head is transmitted to the computer which then recomputes and displays the three-dimensional appearance and the new positions of many objects in the virtual scene. The delay and the update rate depend on several factors and they may vary from scene to scene; they are usually slow enough that the subject has to visually track the jerky motion of whatever object is of his current interest, even if that object is not in motion within the virtual scene (like the orientation grids in the tests described in Section 2). It is then that the eye movements may play a role in sending positional signals that are in conflict with signals from the semicircular canals. Therefore, by keeping the subject's head immobilized in these measurements, we are assuming that the eye movements being measured under these controlled conditions are equivalent (i.e., similar enough for the purpose of these experiments) to those elicited in a VE when head movements trigger the sequence of events that lead to the eye movements described earlier.

## 4.3    RESULTS

Figures 7, 8, and 9 are typical two-dimensional traces of the target trajectory and the subject's eye fixation trajectory, corresponding to runs at 1-, 4-, and 60-Hz position update rates, respectively. They provide a summary representation of the subject's eye movements in a run, but it should be noted that these traces provide information of where but not *when* the eye is fixating with respect to the target. Position as a function of time is necessary for that purpose.

The laboratory apparatus actually provides separate x and y coordinates for the target and the eye position. Given that the target is circumventing a circle at constant speed, its x and y position components are sinusoidal functions of time, and the eye position coordinates are deviations from these sine functions. These coordinates are sampled at 100 Hz, so that for the highest rate of position update tested (60 Hz) there is at least one data point for each different position of the target and a corresponding data point for the position of the subject's eye. Figures 10, 11, and 12 are traces of separate x and y coordinates corresponding to the two-dimensional traces shown in Figures 7, 8, and 9, respectively; they are plotted as functions of time and also display the computed fixation error and eye speed. The formulas used for these computed values are given in Equations (4) and (5). The average values for all update rates obtained from three subjects are listed in Table 2.

$$\text{Fixation error: } E = [ (E_x)^2 + (E_y)^2 ]^{(0.5)} \tag{4}$$

where $E_x$ and $E_y$ are the absolute values of the differences between eye and target x and y coordinates.

$$\text{Eye speed: } S = [ (S_x)^2 + (S_y)^2 ]^{(0.5)} \tag{5}$$

where $S_x$ and $S_y$ are the x and y components of speed of the eye.
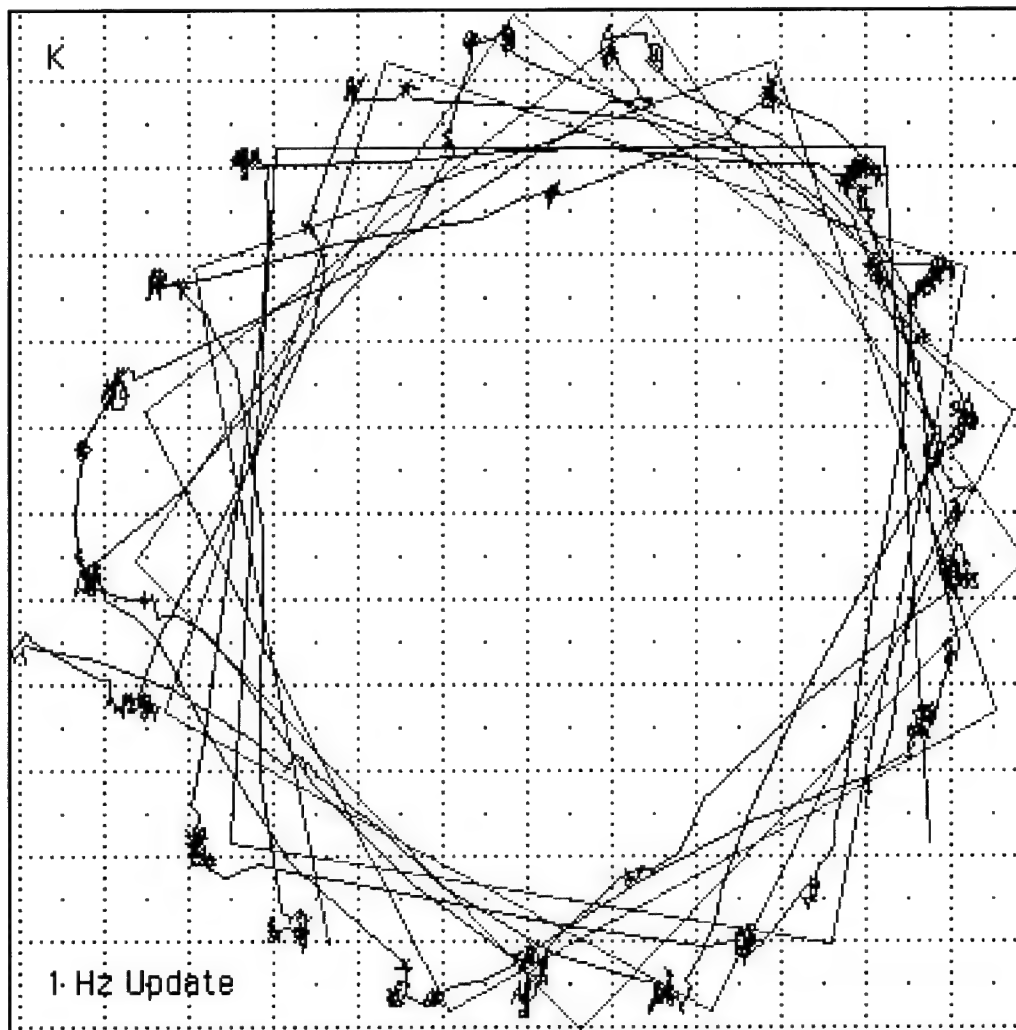
18

Figure 7.    Two-dimensional traces of the test target motion and the corresponding eye
             movements. The red trace corresponds to the test target trajectory and the blue trace
             to the eye movements. These are typical traces obtained with a position update rate
             of 1 Hz.

Figure 8.    Two-dimensional traces of the test target motion and the corresponding eye movements. The red trace corresponds to the test target trajectory and the blue trace to the eye movements. These are typical traces obtained with a position update rate of 4 Hz.
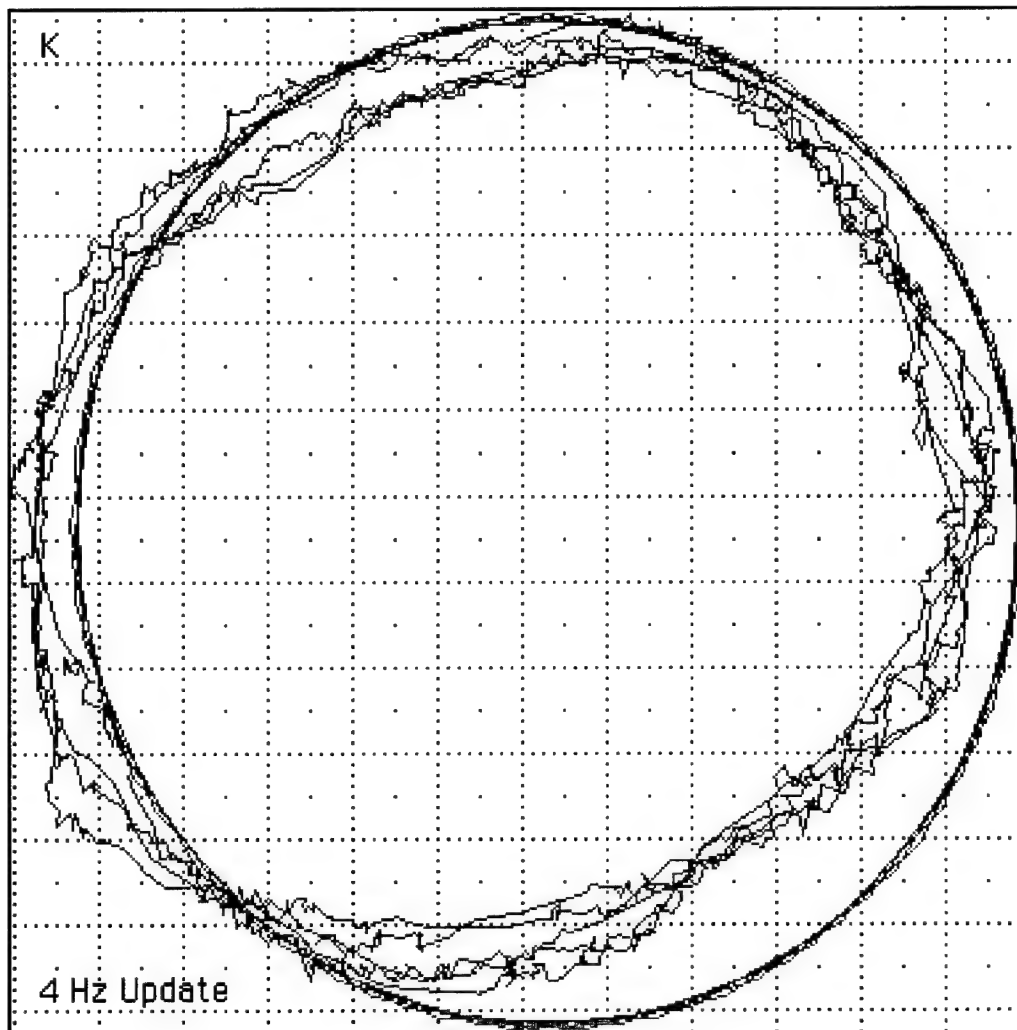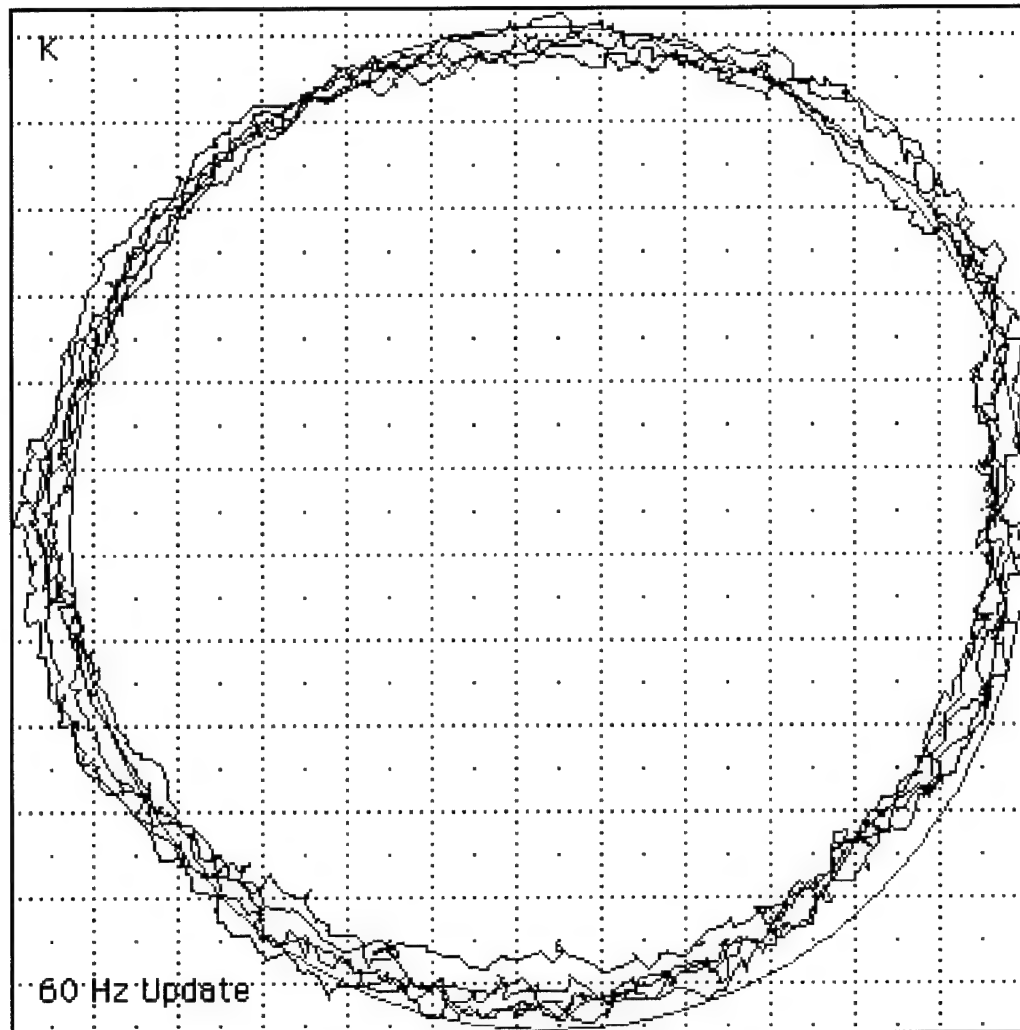
Figure 9.    Two-dimensional traces of the test target motion and the corresponding eye
             movements. The red trace corresponds to the test target trajectory and the blue trace
             to the eye movements. These are typical traces obtained with a position update rate
             of 60 Hz.

21

**Table 2**

**FIXATION ERROR AND EYE SPEED AVERAGE VALUES FOR THREE SUBJECTS**

The left column corresponds to the position update rate in hertz. Subjects are represented by initials K, J, and C. Values for fixation error (err) and eye speed (spd) are in arbitrary units and are averages across all experimental runs and all position update cycles for the update rate given in the first column. The two rightmost columns are summary averages across subjects.

| Rate | K err | K spd | J err | J spd | C err | C spd | avg err | avg spd |
|------|-------|-------|-------|-------|-------|-------|---------|---------|
| 1    | 1.026 | 1.084 | 1.287 | 1.114 | 1.097 | 1.188 | 1.129 | 1.137 |
| 2    | 0.883 | 1.034 | 1.156 | 1.053 | 1.047 | 1.070 | 1.052 | 1.029 |
| 4    | 0.548 | 0.991 | 0.694 | 0.998 | 0.812 | 1.021 | 1.003 | 0.685 |
| 6    | 0.564 | 0.935 | 0.727 | 1.008 | 0.433 | 0.981 | 0.975 | 0.575 |
| 7.5  | 0.498 | 0.953 | 0.628 | 1.054 | 0.363 | 1.000 | 1.002 | 0.496 |
| 8.57 | 0.390 | 0.986 | 0.630 | 1.044 | 0.315 | 1.009 | 1.013 | 0.445 |
| 10   | 0.318 | 0.967 | 0.511 | 0.995 | 0.280 | 1.011 | 0.991 | 0.369 |
| 12   | 0.303 | 0.989 | 0.368 | 0.998 | 0.274 | 0.967 | 0.985 | 0.315 |
| 15   | 0.308 | 1.003 | 0.446 | 0.988 | 0.332 | 1.000 | 0.997 | 0.362 |
| 20   | 0.313 | 0.977 | 0.396 | 0.995 | 0.279 | 0.994 | 0.989 | 0.329 |
| 30   | 0.314 | 0.982 | 0.402 | 1.023 | 0.250 | 0.968 | 0.991 | 0.322 |
| 60   | 0.217 | 1.000 | 0.364 | 1.000 | 0.253 | 1.000 | 1.000 | 0.278 |
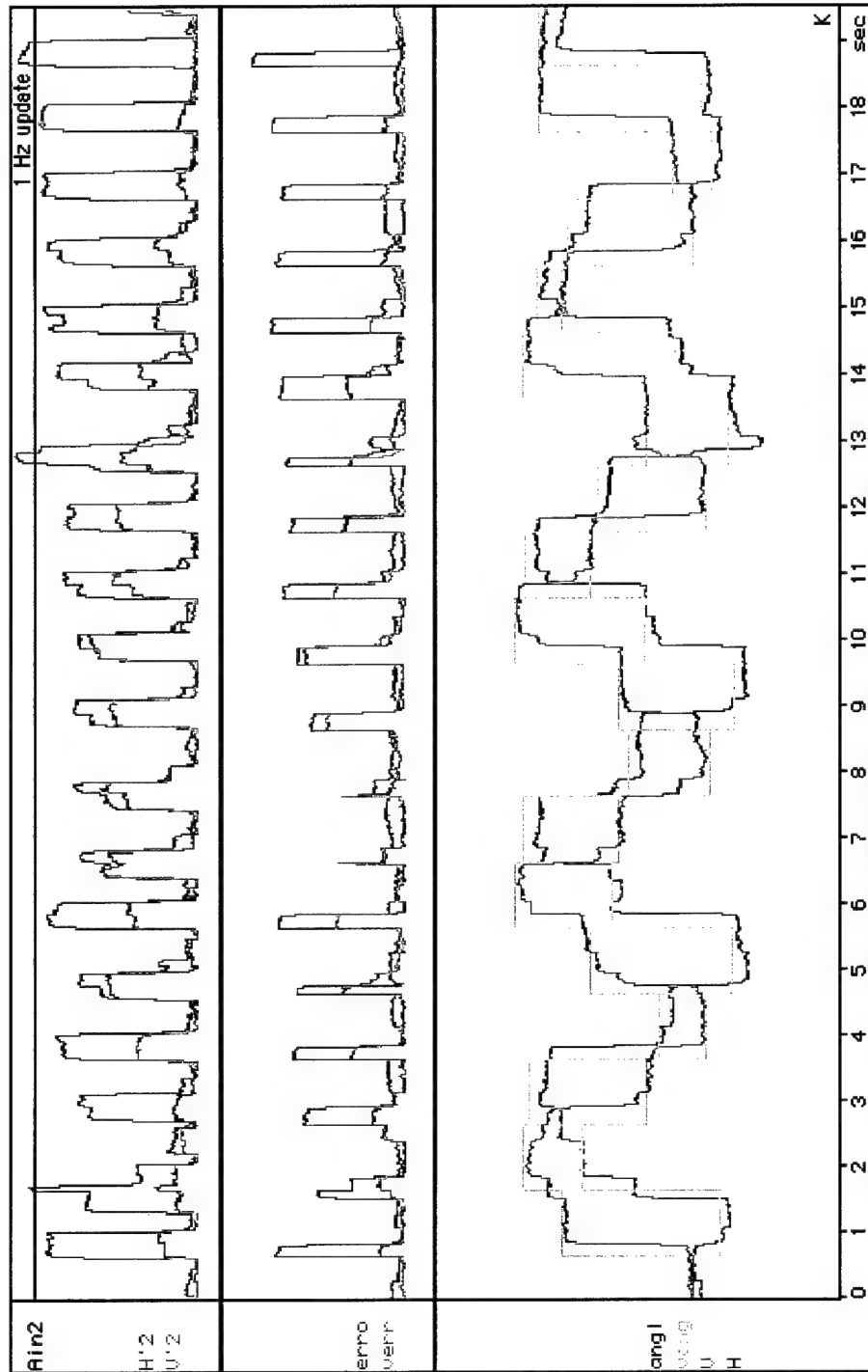
Figure 10. Separate x and y components of motion with 1-Hz update rate.
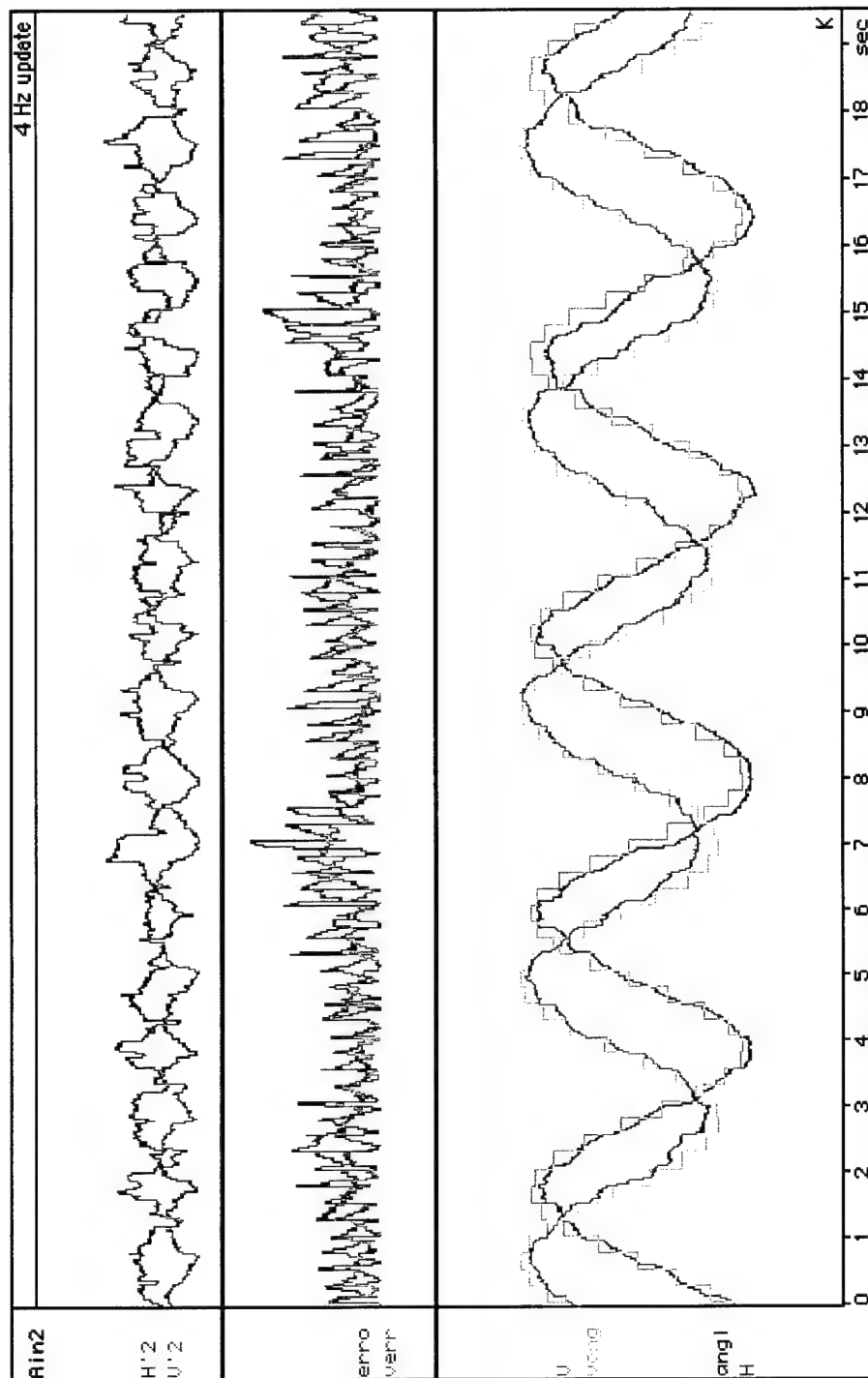
Bottom panel: position values. Middle panel: fixation error values. Top panel: eye speed values. Target x: blue, Target y: yellow, Eye x: red, Eye y: green.

Figure 11. Separate x and y components of motion with 4-Hz update rate.

Bottom panel: position values. Middle panel: fixation error values. Top panel: eye speed values. Target x: blue, Target y: yellow, Eye x: red, Eye y: green.

Figure 12. Separate x and y components of motion with 60-Hz update rate.

Bottom panel: position values. Middle panel: fixation error values. Top panel: eye speed values. Target x: blue, Target y: yellow, Eye x: red, Eye y: green.

Figures 13 through 16 show that the range between 5 and 15 Hz of position update rate seems to have some special characteristics regarding eye movements. Both fixation error and eye speed curves show more instability in that region. That is, there is more variability in error and speed from update to update in position, as well as more variability across subjects. In general, for low update rates, below 5 Hz, error is high and eye speed is high. This is probably due to the predominance of saccadic eye movements that are more inaccurate for visual tracking, thereby producing larger fixation errors and more additional displacements for corrections that bring about higher speeds. For high update rates, 15 Hz and higher, eye speed values decrease asymptotically to a value close to the target's constant speed. This is probably due to the predominance of smooth-pursuit eye movements that are more accurate, produce smaller fixation errors, and require fewer additional displacements for correction. Within the 5 to 15 Hz range, a more balanced but rapidly alternating combination of these two types of eye movements is probably producing the up-and-down variability of speed and fixation error data.
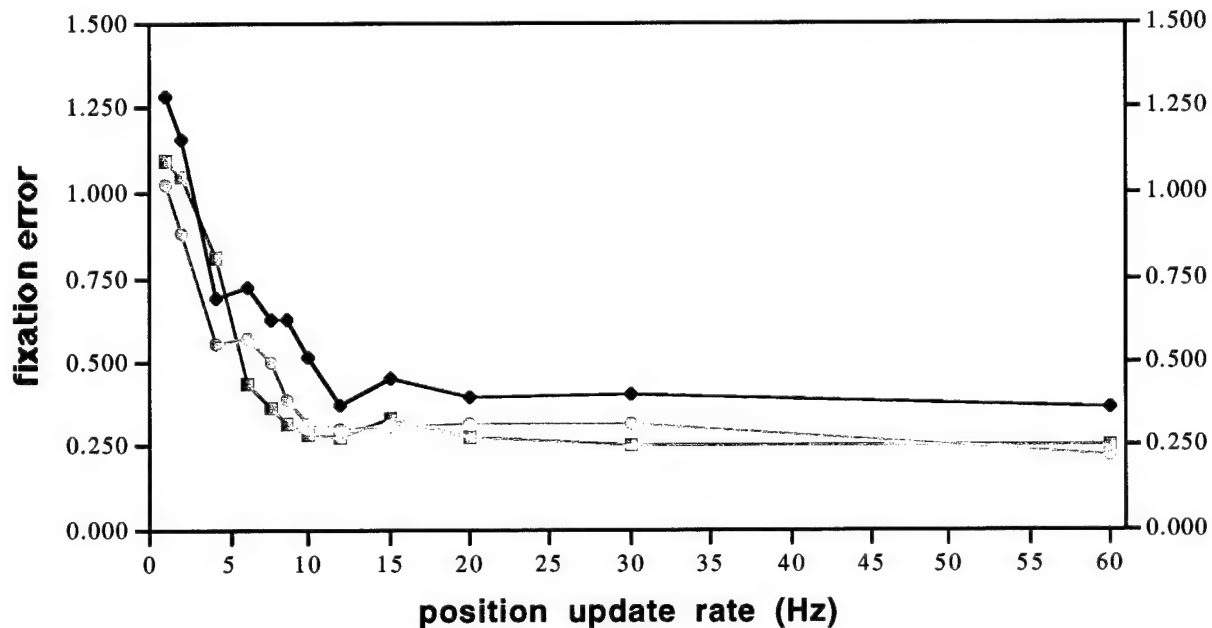


Figure 13.   Individual fixation error values as a function of position update rates. The three sets of data are from three subjects.
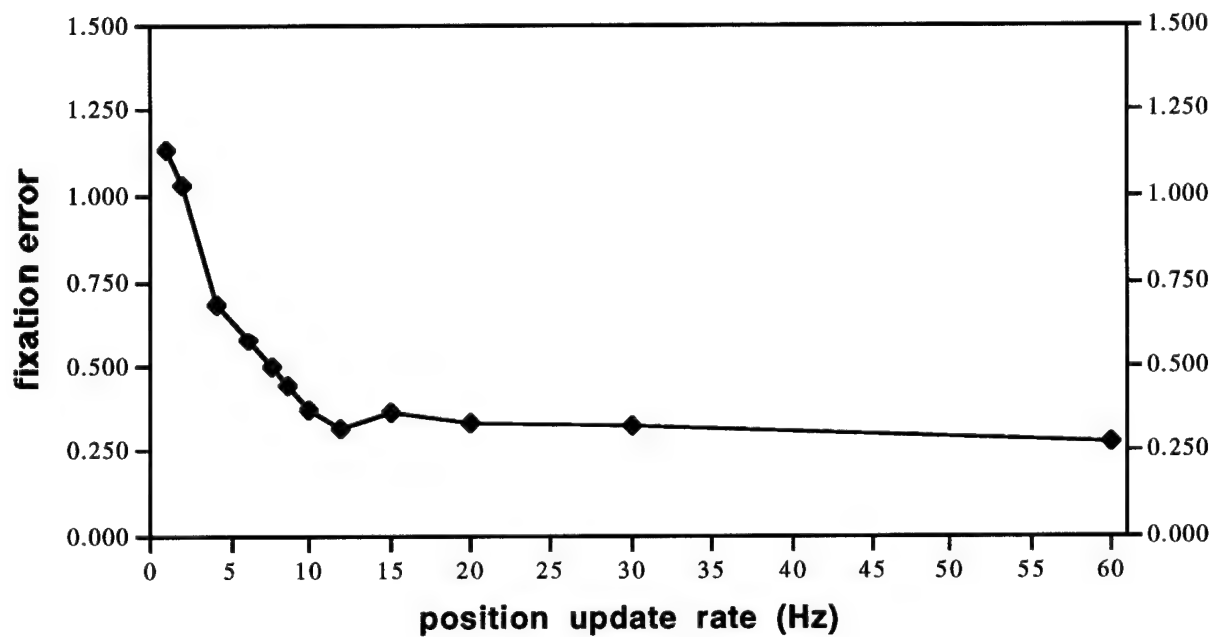
26

Figure 14. Summary fixation error values. Each point in this plot is the average of the three corresponding values of Figure 13.
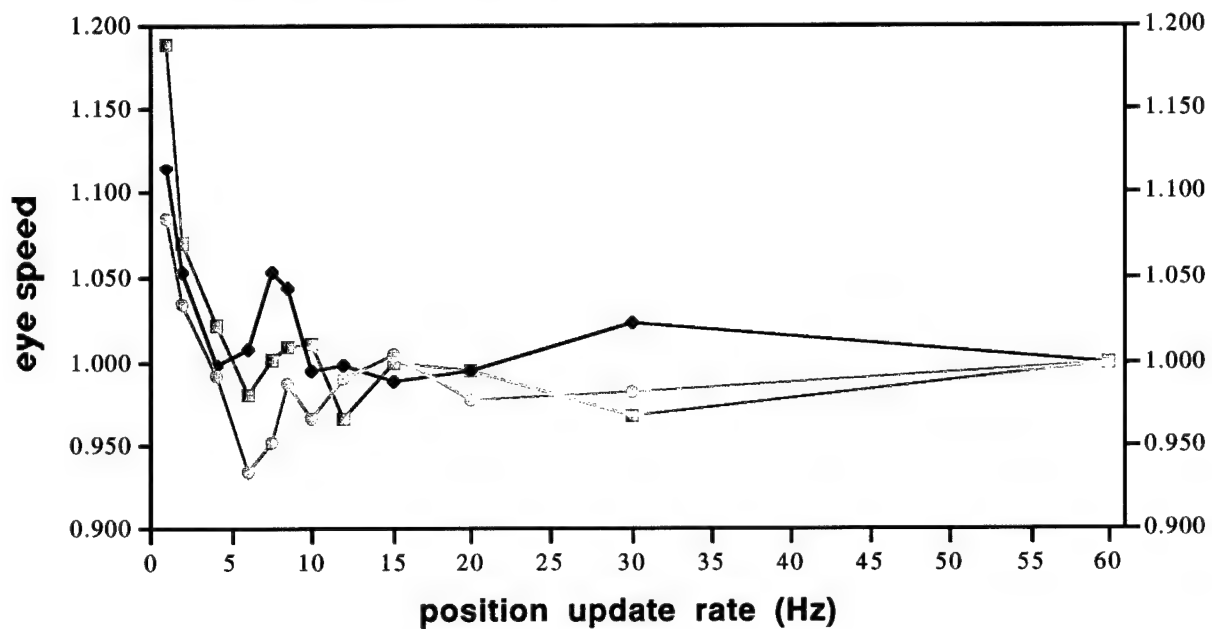


Figure 15. Individual eye speed values as a function of position update rates. The three sets of data are from three subjects.
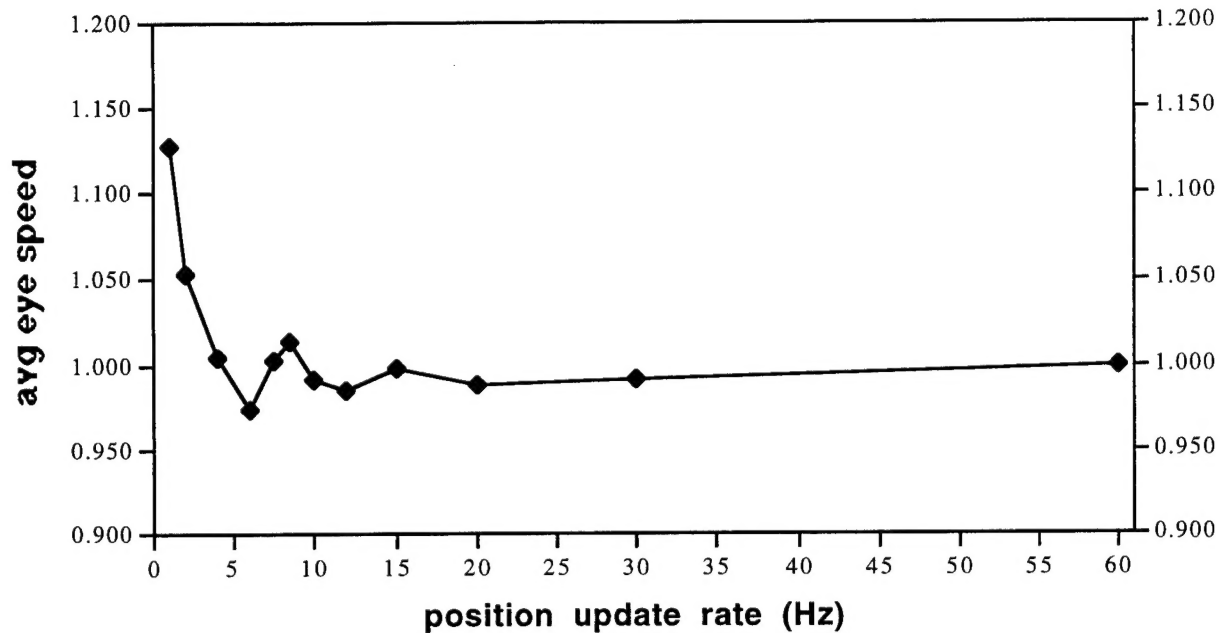
27

Figure 16.   Summary eye speed values. Each point in this plot is the average of the three corresponding values of Figure 15.

An interesting speculation indicated by these results is that, because of their uncommon occurrence in everyday visual tasks, rapidly alternating saccadic and smooth-pursuit eye movements (arising in the 5- to 15-Hz range of position update rate) produce unusual neural signals that are an important component contributing to trigger psychophysiologically unpleasant sensations like vertigo and motion sickness at higher centers of neural processing, probably in combination with conflicting neural signals from the semicircular canals.

Independently of the nature of underlying neural processes, a practical guideline in the design of virtual environments derived from these tests is to avoid position update rates within the 5- to 15-Hz range. For example, in a system where three-dimensional image rendering constraints and/or bandwidth limitations produce position update rates within that range, it might be better to force a reduction in the position update rate. Perceived immersion would then be reduced and optical flow of the scene and objects in motion would be more discontinuous and "jerky", but that measure could avoid detrimental psychophysiological effects in the user.

Another practical guideline, this one related to the use of VE systems for personnel training, is based on the fact that, at least for a few years, many VEs will be delivering position update rates within the problematic range. In that case it should be important to train the VE user to avoid the apparently detrimental combination of the two types of eye movements by controlling the involuntary reflexes triggered by trying to visually fixate a target in jerky motion. That would require the user to learn how to avoid the rapid alternation of saccadic and smooth-pursuit eye movements and voluntarily restrict his eye movements to just smooth pursuit when the optical flow of the VE is jerky and discontinuous.

28

# 5. CONCLUSIONS

The summary results of this project are

- Under the same conditions, target detection performance is better with a larger field of view in the VE. This is a result consistent with many measurements of task performance within real environments.

- Unlike results obtained within real environments, three-dimensional visual orientation grids in a VE can be detrimental for target detection tasks under certain conditions.

- Adding cues to a three-dimensional orientation grid in a VE thereby providing absolute direction of orientation (north, south, east, west) does not improve target detection performance.

- When an acoustic compass is added to relative or absolute visual orientation grids in a VE, target detection performance is reduced. However, adding an acoustic cue attached to the target object in a VE significantly improves target detection performance.

- Update rates of the position of a moving target in VEs are linked to saccadic and smooth-pursuit eye movements. Below 5 Hz, eye movements are predominantly saccadic; above 15 Hz, eye movements are predominantly smooth-pursuit. These two types of eye movements rapidly alternate for intermediate update rates as measured by dynamic fixation error and average eye speed.

Summary guidelines for VE designers derived from these results are

- Improving resolution in head-mounted displays by reducing field of view should be avoided because it severely reduces the user's performance of visual tasks.

- When a designer is planning to increase the field of view of a VE, not only should he take measures to maintain spatial resolution, but also consider a corresponding increase in image update rate to avoid reductions in visual performance. Image update rate is more important for large fields of view, 40 degrees and larger, where, according to the data of this study, an image update rate of the order of 16 Hz is linked to considerable reduction in visual performance.

- The use of acoustic cues similar to an acoustic compass is not recommended for virtual environments because it does not help spatial orientation and introduces unnecessary distraction. However, it is of great benefit to attach an acoustic cue to a virtual object that changes spatial location within the VE, because it increases the user's efficiency to search, find, and identify the object.

29

- It is recommended to avoid image update rates within the 5- to 15-Hz range because they are a factor of negative psychophysiological effects in the user, especially with fields of view 40 degrees and larger.

- For VEs delivering position update rates within 5 to 15 Hz, it is recommended to train the user to avoid the rapid alternation of saccadic and smooth-pursuit eye movements and voluntarily restrict his eye movements to just smooth pursuit when the optical flow of the VE is jerky and discontinuous.

# 6. REFERENCES

1.  Boff, K. R., Kaufman, L., and Thomas, J. P. (Eds.). *Handbook of Perception and Human Performance*, Vol. I, Sensory process and perception. New York. Wiley. 1986.

2.  Boff, K. R., and Lincoln, J. E. *Engineering Data Compendium—Human Perception and Performance*, Vol. III. 1842–1844. Wright Patterson, Ohio: AAMRL (publisher). 1988.

3.  Sheehy, J. B., and Wilkinson, M. Depth perception after prolonged usage of night vision goggles. *Aviation, Space, and Environmental Medicine*. June-89, 573-579. 1989.

4.  Mon-Williams, M., Wann, J. P., and Rushton, S. Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics*, 13-4, 247-298. 1993.

5.  Durlach, N., and Maver, A. *Virtual Reality: Scientific and Technical Challenges*. Washington, D.C: National Academy Press. 1994.

6.  Wenzel, E. M. Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments*, 1(1), 80-107. 1992

7.  Zee, D. S. The vestibulo-ocular reflex: clinical concepts, in B. L. Zuber (Ed.), *Models of Oculomotor Behavior and Control*, 257–278. CRC Press. 1981.